

**Задачи к экзамену по курсу
«Введение в математическую статистику», 2018 г.**

Решения задач можно сдать в бумажном виде преподавателям, ведущим курс, либо отправить им до 23:59 18 декабря в электронном виде (sashaskrip@gmail.com, klimenko05@mail.ru).

В последнем случае все решения должны быть собраны в **один** pdf-файл, в котором страницы работы должны следовать **в правильном порядке и правильной ориентации**. При большом объёме файла можно разместить его в онлайн-хранилище (Яндекс.Диск, Google Drive и т. п.) и прислать только ссылку. Сканы и фото в других форматах рассматриваться не будут.

Также к работе следует приложить файл с расчётами для задачи 7ж) (программу на каком-либо языке программирования, Excel-таблицу с формулами и т. п.).

Баллы за каждую задачу (или отдельный пункт) указаны после номера. Итоговая оценка будет вычисляться по формуле $O = 0,1S - 1,5$, где S — сумма баллов за задачи. Округление производится до ближайшего целого, 0,5 округляется вверх. Если оценка O выходит за границы интервала $[1, 10]$, она заменяется ближайшей его границей.

Задача 1. Учёный Пудель решил построить статистическую модель потребления собаками воды. Он считает, что потребление (в литрах за день) воды собакой весом 18 кг при температуре 20 градусов по Цельсию описывается распределением на $[0, 1]$ с плотностью $f_\theta(x) = \frac{x^\theta}{\theta+1}$. Пудель выбирает между двумя значениями параметра θ : $\theta = 0$ или $\theta = 1$. При этом рассматриваемая им выборка собак состоит только из него самого.

- а) (7) Построить наиболее мощный критерий при заданной вероятности ошибки первого рода α для проверки гипотезы $H_0 : \theta = 0$ против альтернативы $H_1 : \theta = 1$.
- б) (3) Следует ли Пуделю принять гипотезу H_0 , если $\alpha = 0,05$ и сам он выпил 0,7 литра воды за день?
- в) (3) Найдите p -значение для гипотезы H_0 при указанном результате наблюдений.

Задача 2. Рассмотрим выборку (X_1, X_2, \dots, X_n) из распределения с плотностью

$$f(t, \theta) = \frac{2t}{\theta^2}, \quad t \in [0, \theta].$$

- а) (3) Является ли $\hat{\theta} = \frac{3(X_1 + \dots + X_n)}{2n}$ состоятельной оценкой параметра θ ? Является ли эта оценка несмещённой?
- б) (7) Используя эту оценку, предложите доверительный интервал для θ с уровнем доверия 0,95.
- в) (4) Рассмотрим $M = \max(X_1, \dots, X_n)$. Является ли M несмещённой оценкой для θ ? Асимптотически несмещённой?
- г) (4) Является ли M состоятельной оценкой параметра θ ?
- д) (6) Построить оценку максимального правдоподобия для θ . Является ли она эффективной?

Задача 3 (6). Известно, что количество правильно решённых задач на экзамене по статистике распределено по нормальному закону с неизвестными параметрами — средним μ и дисперсией σ^2 . По результатам первого дня экзамена в 2017 году, в котором участвовали 20 студентов, среднее по выборке оказалось равно 10, а выборочная дисперсия — 15. Найти симметричный доверительный интервал для дисперсии σ^2 и полностью обосновать ответ.

Задача 4. Пусть (X_1, \dots, X_n) — выборка объёма n из геометрического распределения с параметром $1/\theta$, $\theta > 1$, т. е. для него $\mathbb{P}(X_i = k) = \frac{1}{\theta} \left(1 - \frac{1}{\theta}\right)^k$ для $k = 0, 1, 2, \dots$

- а) (6) Найти минимальную достаточную статистику для θ .
- б) (4) Найти оценку максимального правдоподобия для θ . Является ли она несмещённой?
- в) (6) Найти информацию Фишера для выборки. Является ли оценка максимального правдоподобия эффективной?

Задача 5 (7). Пусть (X_1, \dots, X_n) — выборка из распределения F_θ . Пусть $E_\theta X_1 = \theta$. Доказать, что вся выборка (X_1, \dots, X_n) не является полной статистикой для параметра θ .

Задача 6 (20). В начале XX века в Великобритании исследовался вопрос о связи глухонемоты с полом. Общая численность населения составляла 15 729 000 мужчин и 16 799 000 женщин, при этом 3 497 мужчин и 3 072 женщины были зарегистрированы как глухонемые от рождения. Предложите критерий для проверки гипотезы о том, что глухонемота связана с полом. Следует ли принять такую гипотезу?

Задача 7. Аналоговый генератор случайных чисел меняет каждую секунду напряжение на своём выходе, при этом получаемые значения напряжения (X_1, \dots, X_n) независимы и равномерно распределены на отрезке $[a\theta, b\theta]$, где $a < 0 < b$ — известные числа, а $\theta > 0$ — коэффициент усиления встроенного усилителя, шкала регулировки которого стёрлась.

- а) (4) Для любых значений $a < 0 < b$ постройте оценку для параметра θ по методу моментов. Является ли она несмещённой?
- б) (4) Найдите достаточную статистику в этой задаче.
- в) (8) Докажите, что построенная статистика полна.
- г) (3) Постройте наилучшую несмещённую оценку для параметра θ .
- д) (5) Выполнено ли для оценки из предыдущего пункта неравенство Крамера—Рао? Если нет, то почему?
- е) (7) Постройте симметричную интервальную оценку для параметра θ с заданным уровнем доверия α .
- ж) (30) Вычислите полученные точечную и интервальную оценку ($\alpha = 90\%$) на предложенном индивидуальном наборе данных.